DERWENT-ACC-NO: 1998-403328

DERWENT-WEEK: 199835

TITLE: Communication method for **computer network** - involves transmitting **data** using remote memory **write-in** function based on position **information** notified by receiving side **computer in packet** assigned in transmitting and receiving **buffer**

PATENT-ASSIGNEE: HITACHI LTD[HITA]

PRIORITY-DATA: 1996JP-0331606 (November 27, 1996)

PATENT-FAMILY:

PUB-NO	PUB-DATE	LANGUAGE	PAGES	MAIN-IPC
JP 10161955 A	June 19, 1998	N/A	012	G06F 013/00

APPLICATION-DATA:

PUB-NO	APPL-DESCRIPTOR	APPL-NO	APPL-DATE
JP 10161955A	N/A	1996JP-0331606	November 27, 1996

INT-CL (IPC): G06F013/00, G06F015/163

ABSTRACTED-PUB-NO: JP 10161955A

BASIC-ABSTRACT:

The method involves initializing communication between multiple computers (101A-101N) based on user's program. A predetermined number of packets are assigned to transmitting and receiving buffer of the remote memory transmitting and receiving area (110) of each computer. During initialization, a computer at the receiving side detects data arrival from a transmitting computer.

After detection, the receiving computer assigns a packet in transmitting and receiving buffer for reception of data to be communicated and then notifies position information of the assigned packet using remote memory write-in function. Then, the transmitting side computer transmits data using remote memory data write-in function based on position information notified by the data receiving computer.

ADVANTAGE - Performs data forwarding at high speed.

CHOSEN-DRAWING: Dwg.1/11

TITLE-TERMS: COMMUNICATE METHOD **COMPUTER NETWORK** TRANSMIT **DATA** REMOTE MEMORY **WRITING** FUNCTION BASED POSITION **INFORMATION** NOTIFICATION RECEIVE SIDE **COMPUTER PACKET** ASSIGN TRANSMIT RECEIVE **BUFFER**

DERWENT-CLASS: T01

EPI-CODES: T01-H07C5;

SECONDARY-ACC-NO:

Non-CPI Secondary Accession Numbers: N1998-314314

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号

特開平10－161955

(43)公開日　平成10年(1998)６月19日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ０６Ｆ　13/00	３５３	Ｇ０６Ｆ　13/00	３５３Ｑ
15/163		15/16	３１０Ｆ

審査請求　未請求　請求項の数４　ＦＤ　（全 12 頁）

(21)出願番号　特願平8－331606

(22)出願日　平成８年(1996)11月27日

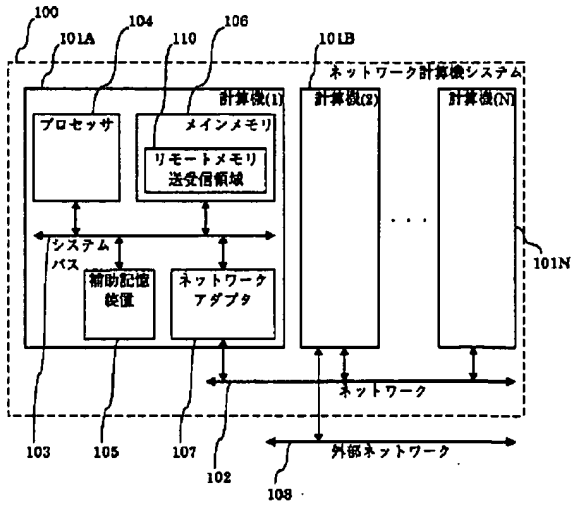
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　吉澤　聡
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内
(72)発明者　岩月　秀樹
神奈川県横浜市戸塚区戸塚町5030番地　株式会社日立製作所ソフトウェア開発本部内
(72)発明者　佐川　暢俊
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内
(74)代理人　弁理士　笹岡　茂　（外１名）

最終頁に続く

(54)【発明の名称】　ネットワーク通信方法

(57)【要約】
【課題】　ネットワーク計算機システムに於いて、リモートメモリ書き込み機構を用いてメッセージ転送を高速かつ安全に行うことにある。
【解決手段】　リモートメモリ書き込み機能を有するネットワーク計算機システムにおいて、各計算機101A、101B、…は、利用者プログラムからの通信初期化要求を契機として、前記リモートメモリ書き込み機能で使用する送受信バッファとして、予め定められた数のフラグメントを有する送受信バッファをリモートメモリ送受信領域110に確保し、受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対し次の通信データの受信用としての送受信バッファ内のフラグメントを割り当て、該フラグメントの位置情報を送信側計算機に対しリモートメモリ書き込み機能を用いて通知し、送信側計算機は、通知された位置情報に基づいて、リモートメモリデータ書き込み機能を用いてデータを送信する。

図１



1

【特許請求の範囲】

【請求項１】　複数の計算機がネットワーク結合され、任意の計算機間でデータ通信を行う際には、データの送信側計算機が、受信側計算機の主記憶装置上の特定位置を指定して通信データを書き込むリモートメモリ書き込み機能を有するネットワーク計算機システムにおけるネットワーク通信方法であって、

各計算機は、利用者プログラムからの通信初期化要求を契機として、前記リモートメモリ書き込み機能で使用する送受信バッファとして、予め定められた数のフラグメントを有する送受信バッファを確保し、

受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対して次の通信データの受信用としての前記送受信バッファ内のフラグメントを指定して割り当て、そのフラグメントの位置情報を送信側計算機に対して前記リモートメモリ書き込み機能を用いて通知し、

送信側計算機は、該受信側計算機により通知された位置情報に基づいて、前記リモートメモリデータ書き込み機能を用いてデータを送信することを特徴とするネットワーク通信方法。

【請求項２】　請求項１記載のネットワーク通信方法において、

前記受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対して次の通信データの受信用としての前記送受信バッファ内のフラグメントを一度に複数個指定して割り当て、その複数個のフラグメントの位置情報を送信側計算機に対して前記リモートメモリ書き込み機能を用いて一度に通知することを特徴とするネットワーク通信方法。

【請求項３】　請求項２記載のネットワーク通信方法において、

前記受信側計算機は、前記一度に複数個指定して割り当てる送受信バッファのフラグメントが主記憶装置上で連続領域となるように割り当てることを特徴とするネットワーク通信方法。

【請求項４】　請求項２記載のネットワーク通信方法において、

前記受信側計算機は、前記送信側計算機に対して既に指定して割り当てた受信用バッファ内のフラグメントのうち、使用済みになっていない使用中フラグメントの数を管理し、該使用中フラグメントの数が予め定められた閾値よりも小さくなった際、該使用中フラグメントの数が該閾値よりも大きい予め定められた値となるように、複数の該フラグメントを受信用に指定して割り当てることを特徴とするネットワーク通信方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は複数の計算機をネットワークによって結合した並列計算機等に代表されるネットワーク計算機システムに於ける計算機間のデータ通信に係わり、特にメッセージパッシングの高速性とデータの安全性の確保に関する。

2

【０００２】

【従来の技術】本発明に関連する技術として「オペレーティングシステムの設計II，ＸＩＮＵによるインターネットワークの構築：ダグラス・カマー著，村井純監修，歌代和正・酒匂順子訳，啓学出版，１９９１年１１月」の２１頁から２４頁記載の技術がある。この従来例では送信側計算機で送信データを複数のパケットに分割し、パケット単位でネットワークに送出する。受信側計算機では、パケットを受信する毎にオペレーティング・システムに対して、割り込み処理を用いてパケットの到着を通知する。通知を受けたオペレーティング・システムは受信したパケットの先頭部分に格納された情報を参照し、到着したパケットが受信データの最終パケットであるか否かを判定する。最終パケットであった場合、通信データの受信側のユーザプログラムが受信待ち状態にあれば割り込みを用いて通信データの到着をＯＳが該ユーザプログラムに通知し、また該ユーザプログラムが受信待ち状態に無い場合には、該ユーザプログラムから受信要求が出される迄、ＯＳで通信データを保持しておき、該ユーザプログラムから受信要求が出された時にそのデータをＯＳが該ユーザプログラムに渡す処理を行う。

【０００３】計算機をネットワークによって結合した並列計算機等に代表されるネットワーク計算機システムに於いては、上記の従来技術に加えて、通信処理速度を向上するための技術として、リモートメモリ書き込み機構がある。リモートメモリ書き込み機構では、各計算機は相手計算機のオペレーティング・システムの介入無しに、即ち、割り込み処理や、パケットのバッファリング処理を行わずに、直接相手計算機内の特定メモリ領域へのデータ転送が可能である。

【０００４】リモートメモリ書き込みを行うことのできる特定メモリ領域は、リモートメモリ送受信領域と呼ばれる。リモートメモリ書き込み機構を有するネットワーク計算機システムに於いて、システム内の異なる計算機間でメッセージパッシングによってデータの送受信を行う際の従来手法を図２を用いて説明する。

【０００５】各計算機１０１Ａ，１０１Ｂ間のネットワーク１０２を通じたデータ転送は、リモートメモリ送受信領域１１０Ａ，１１０Ｂ間で行われ、各計算機はリモートメモリ書き込み機構を用いて書き込みを行う相手計算機の受信バッファ位置情報２２２（アドレス）を保持する。先ず送信関数がコールされると、送信側のユーザプログラム中の送信データ領域２０１からリモートメモリ送受信領域１１０Ａ内の送信バッファ２１１にデータがコピーされる。次にそこから受信側のリモートメモリ送受信領域１１０Ｂ内の受信バッファ２２１に、受信バッファ位置情報２２２を指定してリモートメモリ書き込

み機構を起動することにより、データを転送する。最後に受信関数がコールされて、受信バッファ２２１から受信側のユーザプログラムの受信データ領域２０２にデータがコピーされて、メッセージの受け渡しが完了する。

【０００６】

【発明が解決しようとする課題】上述の様に、リモートメモリ書き込み機構では、相手計算機のオペレーティング・システムの介入無しに高速にデータの転送を行うため、受信側のリモートメモリ送受信領域の使用状態を確認しないと、データを損失する可能性があり、連続したデータ書き込みを安全に行うことができなかった。更に、リモートメモリ送受信領域内の受信バッファの位置及び容量が固定されているため、受信バッファの容量を超える長さのメッセージを転送する事もできなかった。本発明の目的は、任意の長さのメッセージ転送を、限られた容量のリモートメモリ送受信領域を用いて、高速かつデータの損失を起こす事無く、行う事にある。

【０００７】

【課題を解決するための手段】上記課題を解決するため、本発明は、複数の計算機がネットワーク結合され、任意の計算機間でデータ通信を行う際には、データの送信側計算機が、受信側計算機の主記憶装置上の特定位置を指定して通信データを書き込むリモートメモリ書き込み機能を有するネットワーク計算機システムにおけるネットワーク通信方法であり、各計算機は、利用者プログラムからの通信初期化要求を契機として、前記リモートメモリ書き込み機能で使用する送受信バッファとして、予め定められた数のフラグメントを有する送受信バッファを確保し、受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対して次の通信データの受信用としての前記送受信バッファ内のフラグメントを指定して割り当て、そのフラグメントの位置情報を送信側計算機に対して前記リモートメモリ書き込み機能を用いて通知し、送信側計算機は、該受信側計算機により通知された位置情報に基づいて、前記リモートメモリデータ書き込み機能を用いてデータを送信するようにしている。

【０００８】また、前記受信側計算機は、初期化時、並びにデータ到着検知後に、送信側計算機に対して次の通信データの受信用としての前記送受信バッファ内のフラグメントを一度に複数個指定して割り当て、その複数個のフラグメントの位置情報を送信側計算機に対して前記リモートメモリ書き込み機能を用いて一度に通知するようにしている。

【０００９】また、前記受信側計算機は、前記一度に複数個指定して割り当てる送受信バッファのフラグメントが主記憶装置上で連続領域となるように割り当てるようにしている。

【００１０】また、前記受信側計算機は、前記送信側計算機に対して既に指定して割り当てた受信用バッファ内

のフラグメントのうち、使用済みになっていない使用中フラグメントの数を管理し、該使用中フラグメントの数が予め定められた閾値よりも小さくなった際、該使用中フラグメントの数が該閾値よりも大きい予め定められた値となるように、複数の該フラグメントを受信用に指定して割り当てるようにしている。

【００１１】

【発明の実施の形態】図１は本発明の一実施例を示す、ネットワーク計算機システム１００のブロック図である。ネットワーク計算機システム１００は複数の計算機１０１をネットワーク１０２で結合した構成を持つ。ネットワーク計算機システム１００の代表例としては、分散メモリ型の並列計算機を挙げることができる。計算機１０１に於いて、１０３は各構成ブロック間の命令及びデータの転送路であるシステムバス、１０４はプログラムを実行するプロセッサ、１０５はプログラムを記述したコード等を格納する磁気ディスク装置等の補助記憶装置、１０６はプログラム実行中にそのコードやデータを格納するメインメモリ、１０７はネットワーク１０２との入出力を制御するネットワークアダプタである。また、１０８はネットワーク計算機システム１００と外部の計算機を接続するための外部ネットワークであり、１００内の一つ以上の計算機１０１が接続されている。外部ネットワーク１０８を通じてネットワーク計算機システム１００、端末装置や大容量ディスク装置（ファイル・サーバ）等が接続される。本発明のネットワーク計算機システム１００に於いては、任意の計算機間でデータ通信を行う際に、データ送信側の計算機が、受信側計算機のメインメモリ上の特定位置を指定して、通信データを書き込むことのできる、リモートメモリ書き込み機構を有する。リモートメモリ書き込み機構で使用するメインメモリ１０６上の領域は、１１０のリモートメモリ送受信領域である。尚、図1では省略しているが、計算機１０１では、表示装置、キーボードやマウス等の入出力装置がシステムバス１０３に接続されていても良い。

【００１２】図３、図４は、それぞれ送信側計算機１０１Ａ、受信側計算機１０１Ｂの、メインメモリ１０６Ａ、１０６Ｂ上に確保される領域の一実施例を示す図である。図３に示した送信側計算機１０１Ａのメインメモリ１０６Ａの実施例に於いては、２０１はユーザプログラムが送信するメッセージを格納した送信データ領域である。本発明では、リモートメモリ書き込み機構を用いたデータの送受信処理を、予め定められたデータ長で行い、この送受信処理の単位を「フラグメント」と呼ぶ。即ち、送信データ領域２０１の送信処理はフラグメント長毎に行い、２０１－１、２０１－２、…のフラグメントに分割して行う。３０１は、通信相手に対して既に送信したフラグメントの個数を格納する送信済みフラグメント通番情報であり、値「０」に初期化する。送信済みフラグメント通番情報３０１は通信相手毎に夫々別々に

用意される。リモートメモリ送受信領域１１０Ａには、通信相手に対応して３１１の受信フラグメント位置情報管理テーブルが設けられる。該テーブルには＃１～＃ｊのエントリが用意される。また、受信フラグメント位置情報管理テーブル３１１は通信相手毎に夫々別々に用意される。また、送受信フラグメント領域３２１Ａが設けられ、複数のフラグメント３２１Ａ－１、３２１Ａ－２、・・・、３２１Ａ－ｋに分割される。送受信フラグメント領域３２１Ａリモートメモリ送受信領域に１つ用意される。ユーザプログラムは、これらの領域は、ユーザプログラムの通信初期化要求等を契機として送受信フラグメント領域３２１Ａ内のいくつかのフラグメントを確保する。ユーザプログラムに対応する受信フラグメント位置情報管理テーブル３１１の各エントリは上記初期化要求時に値「－１」に初期化され、その後確保されたフラグメントのフラグメント番号（＃１～＃ｋ）がテーブル３１１のエントリに格納される。

【００１３】図４に示した受信側計算機１０１Ｂのメインメモリ１０６Ｂの実施例に於いては、２０２はユーザプログラムが受信するメッセージを格納する受信データ領域である。また４０１は、通信相手から既に受信したフラグメントの個数を格納する受信済みフラグメント通番情報、４０２は、通信相手に対して既に通知済みの受信用フラグメントの個数を格納する予約済みフラグメント通番情報である。受信済みフラグメント通番情報４０１、予約済みフラグメント通番情報４０２は通信相手毎に夫々別々に用意される。４０１及び４０２は、それぞれ値「０」に初期化する。リモートメモリ送受信領域１１０Ｂには、通信相手に対して４１１の受信フラグメント予約情報管理テーブルが設けられる。該テーブル４１１は通信相手毎に夫々別々に用意される。本テーブル４１１のエントリ（＃１～＃ｊ）のエントリ数は、受信フラグメント位置情報管理テーブル３１１のエントリ数に一致させる。また、送受信フラグメント領域３２１Ｂも、送信側計算機１０１Ａの場合と同様に設けられる。ユーザプログラムは、ユーザプログラムの通信初期化要求等を契機として送受信フラグメント領域３２１Ｂ内のいくつかのフラグメントを確保する。ユーザプログラムに対応する受信フラグメント予約情報管理テーブル４１１の各エントリは上記通信初期化要求時に値「－１」に初期化され、その後確保されたフラグメントのフラグメント番号（＃１～＃ｋ）がテーブル３１１のエントリに格納される。図５、図６は、それぞれ送信側計算機１０１Ａ、受信側計算機１０１Ｂに於ける本発明の一実施例の構成を示すブロック図である。本実施例にて示す各手段は、プログラムとして記述されプロセッサ１０４にて実行されるが、これら手段の一部、または全ては、例えばネットワークアダプタ１０７上にて実施しても良い。尚、各手段の処理内容については、後でフローチャートにて詳述する。

【００１４】図５に示した送信側計算機１０１Ａの実施例に於いては、５１０は送信データ領域２０１から送受信フラグメント領域３２１内のフラグメント３２１Ａ－ｘにデータをコピーし、ネットワークアダプタ１０７を起動する、データ送信制御処理手段である。該処理手段による処理については図８のフローチャートにより説明する。５３０は送信済みフラグメント通番情報３０１と受信フラグメント位置情報管理テーブル３１１を参照して受信側計算機１０１Ｂの送受信フラグメント３２１Ｂ－ｙの位置情報を割り出す、受信フラグメント位置情報取得制御処理手段である。該処理手段による処理については図７のフローチャートにより説明する。５５０Ａは、送受信フラグメント領域３２１Ａの中で、未使用のフラグメントを管理する、未使用フラグメント管理制御処理手段である。即ち、該手段は、送信データ領域２０１のフラグメントのデータを送受信フラグメント領域３２１Ａにコピーするとき、該領域３２１Ａ内の未使用のフラグメントを与え、該領域３２１Ａ内のフラグメントのデータが受信側に転送されたとき、該データ転送済みのフラグメントの未使用登録を受け、未使用フラグメントの管理をする。

【００１５】図６に示した受信側計算機１０１Ｂの実施例に於いては、６１０は送受信フラグメントへのデータの到着を検知し、そこから受信データ領域２０２にデータをコピーする、データ受信制御処理手段である。該処理手段による処理については図９のフローチャートにより説明する。６３０は、受信済みフラグメント通番情報４０１及び予約済みフラグメント通番情報４０２を参照して、必要に応じて受信フラグメント予約情報管理テーブル４１１を更新し、リモートメモリ書き込み機構を用いてテーブル４１１の内容を送信側計算機１０１Ａの受信フラグメント位置情報管理テーブル３１１に書き込む機能、及び送信側計算機１０１Ａからの次の送信データが書き込まれる送受信フラグメント３２１Ｂ－ｙの位置を割り出す機能を有する、受信フラグメント予約情報取得通知制御処理手段である。該処理手段による処理については図１０のフローチャートにより説明する。５５０Ｂは、送信側計算機１０１Ａの５５０Ａと同様に、未使用フラグメント管理制御処理手段である。即ち、該手段は、送受信フラグメント領域３２１Ｂ内のフラグメントのデータを受信データ領域２０２にコピーしたとき、該領域３２１Ｂ内のフラグメントの未使用登録を受け、受信フラグメント予約情報管理テーブル４１１に該領域３２１Ｂ内のフラグメントを登録する際、登録された未使用フラグメントを与え、未使用フラグメントの管理をする。

【００１６】図７は、データ送信制御処理手段５１０の一実施例を示すフローチャートである。ステップ７０１で先ずデータ送信制御処理手段５１０内にあるカウンタＮを１に初期化する。ステップ７１０では、未使用フラ

グメント管理制御処理手段５５０Ａから、未使用フラグメント３２１Ａ－ｘを取得し、ステップ７２０で、該取得したフラグメント３２１Ａ－ｘに送信データ領域２０１のＮ番目のフラグメントのデータ２０１－Ｎをコピーする。ステップ７３０では、受信フラグメント位置情報取得制御処理手段５３０を実行して、受信側計算機１０１Ｂの受信フラグメントの位置情報を取得する。尚、手段５３０の詳細については図８のフローチャートで説明する。ステップ７４０では、ネットワークアダプタ１０７Ａを起動して、上記フラグメント３２１Ａ－ｘのデータをネットワーク１０２に送出、受信側計算機１０１Ｂのフラグメント３２１Ｂ－ｙに書き込む。ステップ７５０では、未使用フラグメント管理制御処理手段５５０Ａに、上記データ送出の終わったフラグメント３２１Ａ－ｘを未使用として登録する。ステップ７６０でカウンタＮの値に１を加算し、最後にステップ７７０で送信データ領域２０１内の全てのデータの送信が完了したか否かを判定する。送信が完了していれば５１１で、データ送信制御処理５１０を終了し、完了していなければステップ７１０に戻り、ステップ７７０迄を繰り返し実行する。本実施例に於いては、フラグメントを一つづつステップ７４０でネットワーク１０２に送出したが、これは複数個づつ一括して送出することとしても良い。その場合、ステップ７５０でのフラグメントの未使用登録処理も、フラグメント送出後に一括して行う。ネットワーク１０２への送出処理を一括化することにより、ネットワークアダプタ１０７Ａを起動することによるオーバヘッドを軽減することが可能である。

【００１７】図８は、受信フラグメント位置情報取得制御処理手段５３０の一実施例を示すフローチャートであり、図７のステップ７３０にて呼び出される。ステップ８１０で先ず、送信済みフラグメント通番情報３０１を読み込み、値を受信フラグメント位置情報取得制御処理手段５３０内にあるカウンタＲに設定する。ステップ８２０では、カウンタＲに設定された値に１を受信フラグメント位置情報管理テーブル３１１のエントリ数（ｊ）で割った時の余りｉ１を加算する計算をし、その計算値番目のエントリの内容を受信フラグメント位置情報管理テーブル３１１から読み出して、手段５３０内にあるカウンタＤに設定する。例えば、図３の場合、送信済みフラグメント通番情報が「０」で、カウンタＲの値が「０」なら上記計算値は「１」となり、受信フラグメント位置情報管理テーブル３１１における計算値番目のエントリの内容は「＃３」となる。ステップ８３０で、Ｄの値が「－１」に一致した場合には、ステップ８２０を繰り返し実行する。このときＲの値は変化しない。なお、受信フラグメント位置情報管理テーブル３１１の内容は、図１０のフローチャートで説明するように、受信フラグメント予約情報管理テーブル４１１に設定された内容が転送、複写される。「－１」に一致しなかった場合には、ステップ８４０に進み、Ｒに１加算した値を送信済みフラグメント通番情報３０１に設定し、５３１でカウンタＤの値をステップ７３０に対する戻り値として、受信フラグメント位置情報取得制御処理５３０を終了する。

【００１８】図９は、データ受信制御処理手段６１０の一実施例を示すフローチャートである。ステップ９０１で先ず上記手段６１０内のカウンタＮを１に初期化する。ステップ９１０では、受信フラグメント予約情報取得制御処理手段６３０を実行して、送信側計算機１０１Ａが次に書き込む自計算機１０１Ｂ内の受信フラグメント３２１Ｂ－ｙの位置情報を取得する。尚、手段６３０の詳細については図１０のフローチャートで説明する。ステップ９２０では、ステップ９１０で割り出した受信フラグメントにデータが到着しているか否かを調べる。ステップ９２０は、データの到着を検知する迄、繰り返し実行し、データの到着を検知した場合には、ステップ９３０に進む。９３０では、到着したデータを、受信データ領域内のＮ番目のフラグメントに該当する位置にコピーする。ステップ９４０では、未使用フラグメント管理制御処理手段５５０Ｂに、上記コピーの終わったフラグメント３２１Ｂ－ｙを未使用として登録する。なお、該魅しよう登録されたフラグメントに該当する受信フラグメント予約情報管理テーブル内のフラグメントの番号は－１に更新される。

【００１９】ステップ９５０でカウンタＮの値に１を加算し、最後にステップ９６０で受信データ領域２０２内の全てのデータの受信が完了したか否かを判定する。受信が完了していれば６１１で、データ受信制御処理６１０を終了し、完了していなければステップ９１０に戻り、ステップ９６０迄を繰り返し実行する。

【００２０】図１０は、受信フラグメント予約情報取得通知制御処理手段６３０の一実施例を示すフローチャートであり、図９のステップ９１０にて呼び出される。ステップ１０１０で先ず、受信済みフラグメント通番情報４０１を読み込み、値を上記手段６３０内のカウンタＳに設定する。ステップ１０２０では、予約済みフラグメント通番情報４０２を読み込み、値を上記手段６３０内のカウンタＴに設定する。次にステップ１０３０でＴとＳの値の差を計算する。（Ｔ－Ｓ）の値は、送信側計算機１０１Ａに対して通知してあり、かつ未だデータが到着していない受信用フラグメントの数に該当する。

【００２１】（Ｔ－Ｓ）の値が予め定められた閾値ＬＷＭ（Ｌｏｗ　Ｗａｔｅｒ　Ｍａｒｋ）を下回る場合、ステップ１０６０以降に進み、受信フラグメントの予約、及び通知の処理を行う。閾値ＬＷＭを下回らない場合はステップ１０４０に進む。ステップ１０６０では、未使用フラグメント管理制御処理手段５５０Ｂより未使用のフラグメントを取得する。ステップ１０６５では、カウンタＴに設定された値を受信フラグメント予約情報管理

テーブル４１１のエントリ数（ｊ）で割った時の余りに１を加算する計算をし、受信フラグメント予約情報管理テーブル４１１の、該計算で得られた計算値番目のエントリにステップ１０６０で取得したフラグメントの位置情報（例えば、＃３、＃５など）を設定する。ステップ１０７０では、カウンタＴの値を１加算し、再度（Ｔ－Ｓ）を計算する。（Ｔ－Ｓ）の値が、予め定められた閾値ＭＡＸよりも小さい場合には、ステップ１０６０に戻り、ステップ１０７５迄を繰り返し実行する。ここで閾値ＭＡＸの値は、送信側計算機１０１Ａに対して通知してあり、かつ未だデータが到着していない受信用フラグメントの最大許容数に相当する。また閾値ＭＡＸの値は、受信フラグメント予約情報管理テーブル４１１のエントリ数を超えない値である。ステップ１０７５で、（Ｔ－Ｓ）の値が閾値ＭＡＸよりも小さくなくなった場合には、ステップ１０８０に進み、受信フラグメント予約情報管理テーブル４１１を、リモートメモリ書き込み機構を用いて、送信側計算機１０１Ａの受信フラグメント位置情報管理テーブル３１１に書き込む。ステップ１０８５では、カウンタＴの値を、予約済みフラグメント通番情報４０２に設定し、ステップ１０４０に進む。ステップ１０３０で、（Ｔ－Ｓ）の値が閾値ＬＷＭを下回らない場合にも、ステップ１０４０に進む。ステップ１０４０では、カウンタＳに設定された値を受信フラグメント予約情報管理テーブル４１１のエントリ数（ｊ）で割った時の余りに１を加算する計算をし、その計算値番目のエントリの内容を受信フラグメント予約情報管理テーブル４１１から読み出して、上記手段６３０内のカウンタＢに設定する。ステップ１０５０で、カウンタＳの値に１加算して、受信済みフラグメント通番情報４０１に設定する。６３１で、カウンタＢの値をステップ９１０に対する戻り値として、受信フラグメント予約情報取得通知制御処理６３０を終了する。

【００２２】図１１に送受信フラグメント領域３２１内の各フラグメントの使用状態を管理する方式の一実施例を示す。１１１０は送受信フラグメント管理テーブルであり、１１１０の各エントリは、送受信フラグメント領域３２１内の各フラグメント３２１－１、・・・、３２１－ｋに対応する。１１２０は未使用フラグメント管理情報であり、現在未使用のフラグメントに対応する送受信フラグメント管理テーブル１１１０のエントリのチェインの先頭を指し示す。１１３０は使用中フラグメント管理情報であり、現在使用中のフラグメントに対応する送受信フラグメント管理テーブル１１１０のエントリのチェインの先頭を指し示す。送受信フラグメント管理テーブル１１１０の全てのエントリは、必ず上記何れか片方のチェインに属する。尚本実施例では、テーブル１１１０、１１２０共に、チェインの先頭のみを指し示したが、チェインの終端も指し示す様に実施しても良い。

【００２３】

【発明の効果】本発明によれば、複数の計算機をネットワークによって結合した並列計算機等に代表されるネットワーク計算機システムに於いて、送受信バッファの割り当てを、受信側から送信処理に先立ち必要に応じて行うので、リモートメモリ書き込み機構を用いて、任意の長さのメッセージ転送を、限られた容量のリモートメモリ送受信領域で、高速かつデータの損失を起こす事無く、行うことができる。

【図面の簡単な説明】

【図１】本発明の一実施例のネットワーク計算機システムのブロック図である。

【図２】ネットワーク計算機システムに於ける通信制御方式の従来例を示すブロック図である。

【図３】本発明のネットワーク計算機システムに於ける、送信側計算機の一実施例を示す説明図である。

【図４】本発明のネットワーク計算機システムに於ける、受信側計算機の一実施例を示す説明図である。

【図５】本発明のネットワーク計算機システムに於ける、送信側計算機の一実施例の構成を示すブロック図である。

【図６】本発明のネットワーク計算機システムに於ける、受信側計算機の一実施例の構成を示すブロック図である。

【図７】本発明のデータ送信制御処理部の一実施例を示すフローチャートである。

【図８】本発明の受信フラグメント位置情報取得制御処理部の一実施例を示すフローチャートである。

【図９】本発明のデータ受信制御処理部の一実施例を示すフローチャートである。

【図１０】本発明の受信フラグメント予約情報取得通知制御処理部の一実施例を示すフローチャートである。

【図１１】本発明の送受信フラグメント管理情報の一実施例を示すブロック図である。

【符号の説明】

１００　ネットワーク計算機システム
１０１　計算機
１０２　ネットワーク
１０３　システム・バス
１０４　プロセッサ
１０５　補助記憶装置
１０６　メインメモリ
１０７　ネットワークアダプタ
１０８　外部ネットワーク
１１０　リモートメモリ送受信領域
２０１　送信データ領域
２０２　受信データ領域
３０１　送信済みフラグメント通番情報
３１１　受信フラグメント位置情報管理テーブル
３２１　送受信フラグメント領域
４０１　受信済みフラグメント通番情報

１１

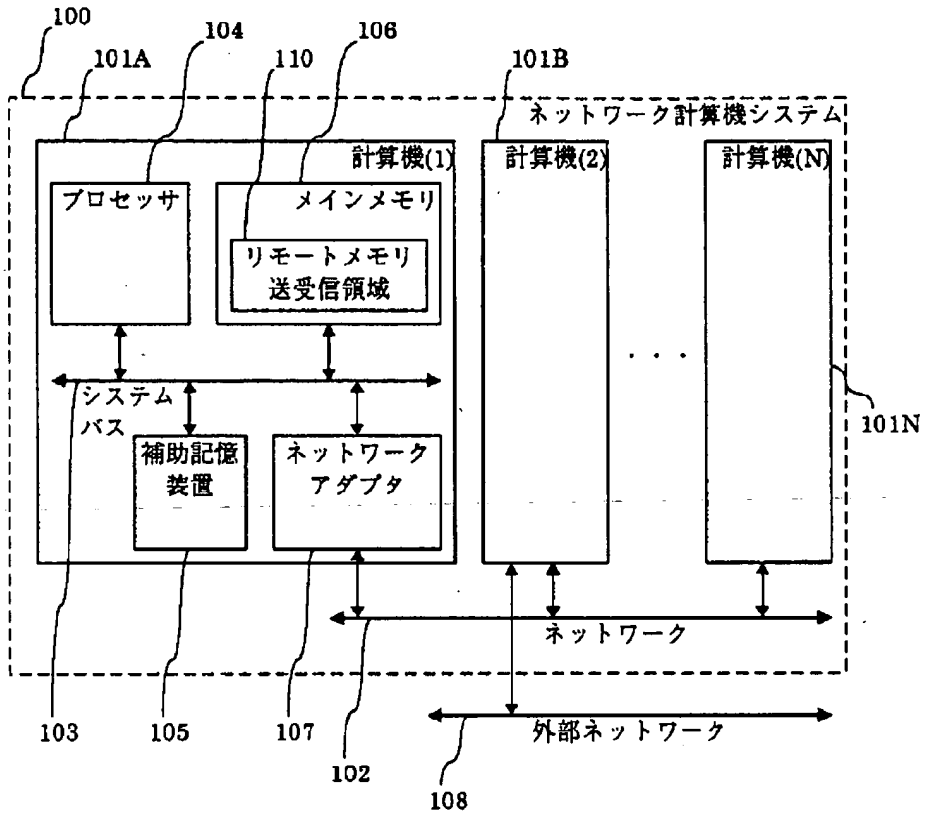
４０２　予約済みフラグメント通番情報
４１１　受信フラグメント予約情報管理テーブル
５１０　データ送信制御処理手段
５３０　受信フラグメント位置情報取得制御処理手段
５５０　未使用フラグメント管理制御処理手段
６１０　データ受信制御処理手段

１２

６３０　受信フラグメント予約情報取得通知制御処理手段
７０１－１０８５　処理ステップ
１１１０　送受信フラグメント管理テーブル
１１２０　未使用フラグメント管理情報
１１３０　使用中フラグメント管理情報

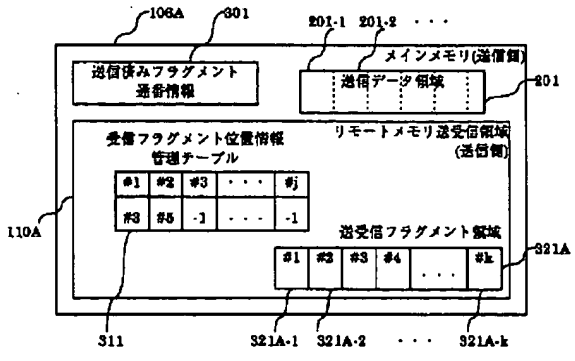
【図１】

図 1



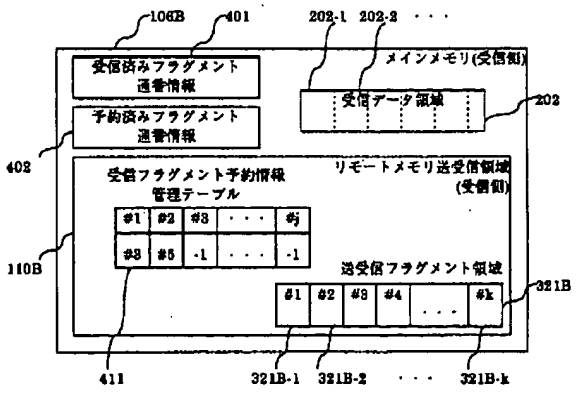
【図３】

図 3



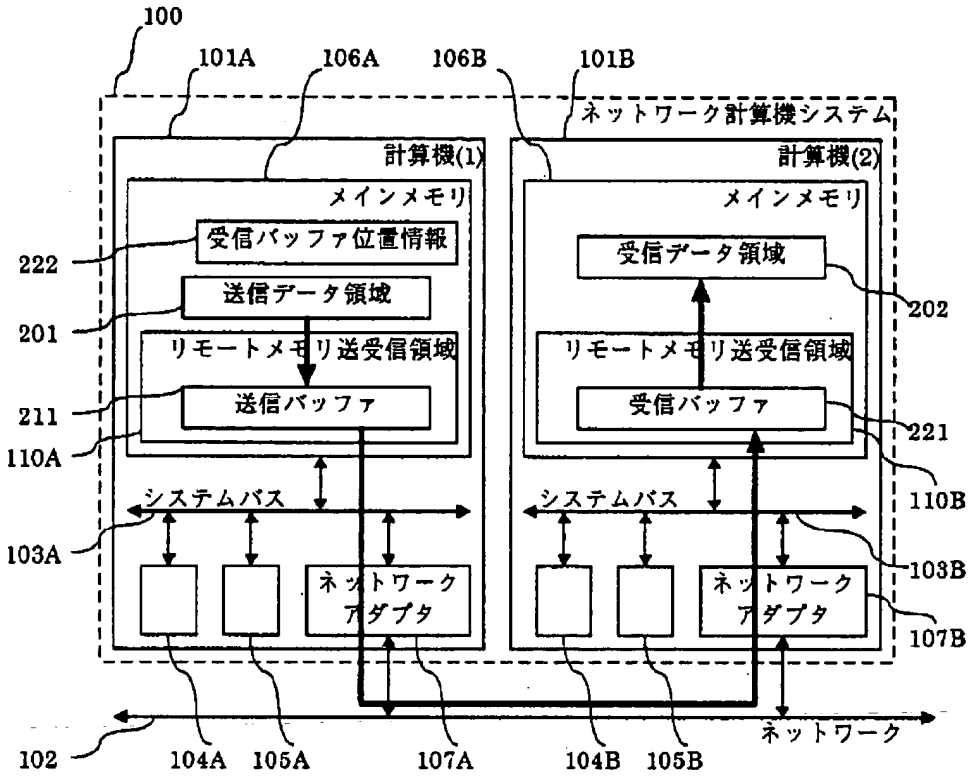
【図４】

図 4



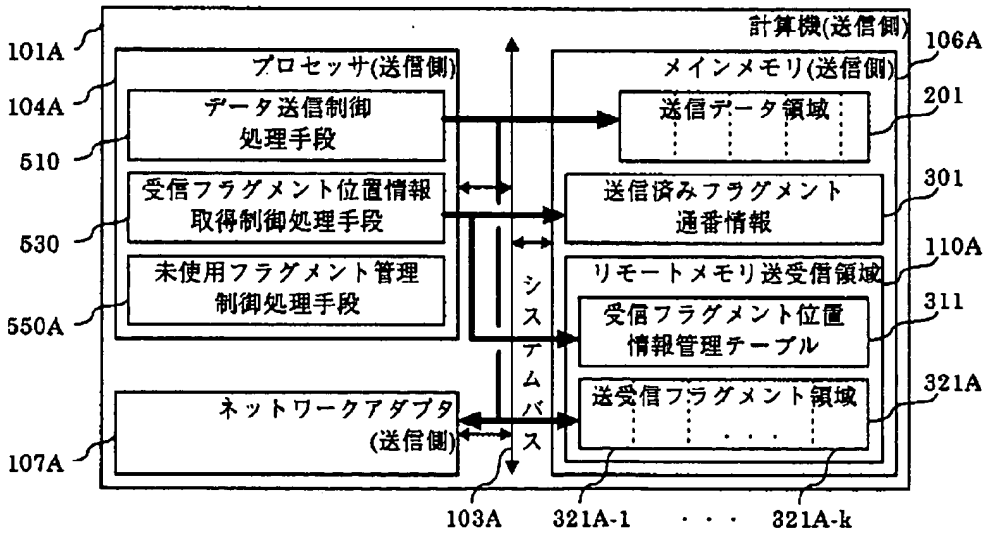
【図2】

図 2


100
101A
106A
106B
101B
ネットワーク計算機システム
計算機(1)
計算機(2)
メインメモリ
メインメモリ
受信バッファ位置情報
222
送信データ領域
201
受信データ領域
202
リモートメモリ送受信領域
リモートメモリ送受信領域
送信バッファ
211
受信バッファ
221
110A
110B
システムバス
システムバス
103A
103B
ネットワークアダプタ
ネットワークアダプタ
107B
ネットワーク
102
104A
105A
107A
104B
105B

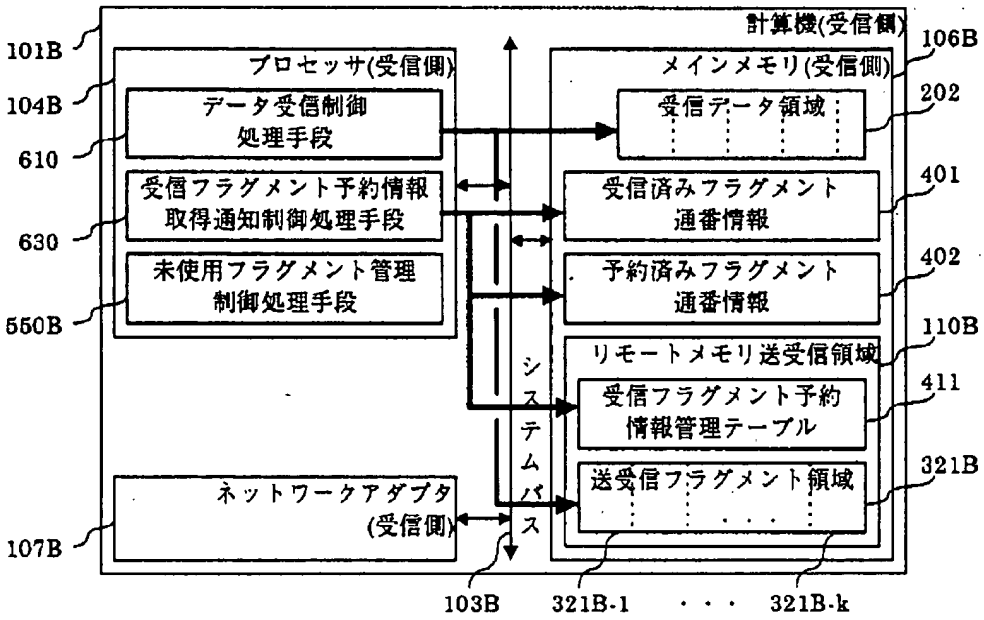
【図5】

図 5


計算機(送信側)
101A
106A
プロセッサ(送信側)
メインメモリ(送信側)
104A
データ送信制御
処理手段
送信データ領域
201
510
受信フラグメント位置情報
取得制御処理手段
送信済みフラグメント
通番情報
301
530
未使用フラグメント管理
制御処理手段
リモートメモリ送受信領域
110A
550A
受信フラグメント位置
情報管理テーブル
311
システムバス
ネットワークアダプタ
(送信側)
送受信フラグメント領域
321A
107A
103A
321A-1
・・・
321A-k

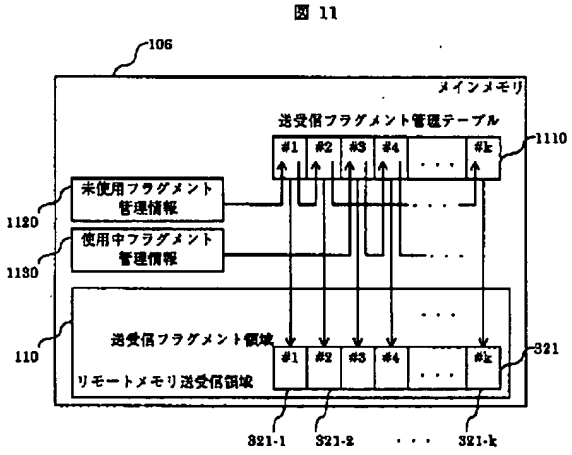
【図6】

図 6


計算機(受信側)
101B
プロセッサ(受信側)
104B
データ受信制御
処理手段
610
受信フラグメント予約情報
取得通知制御処理手段
630
未使用フラグメント管理
制御処理手段
550B
ネットワークアダプタ
(受信側)
107B
106B
メインメモリ(受信側)
202
受信データ領域
401
受信済みフラグメント
通番情報
402
予約済みフラグメント
通番情報
110B
リモートメモリ送受信領域
411
受信フラグメント予約
情報管理テーブル
321B
送受信フラグメント領域
システムバス
103B
321B-1
・・・
321B-k

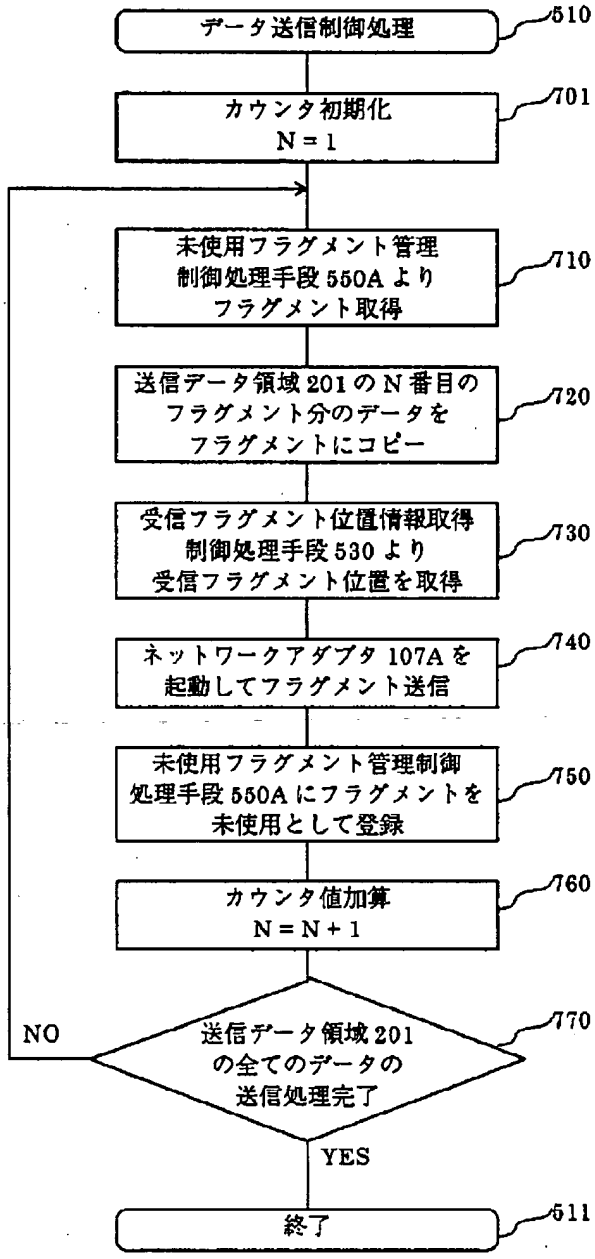
【図11】

図 11


106
メインメモリ
送受信フラグメント管理テーブル
#1
#2
#3
#4
#k
1110
未使用フラグメント
管理情報
1120
使用中フラグメント
管理情報
1130
110
送受信フラグメント領域
リモートメモリ送受信領域
321
321-1
321-2
321-k

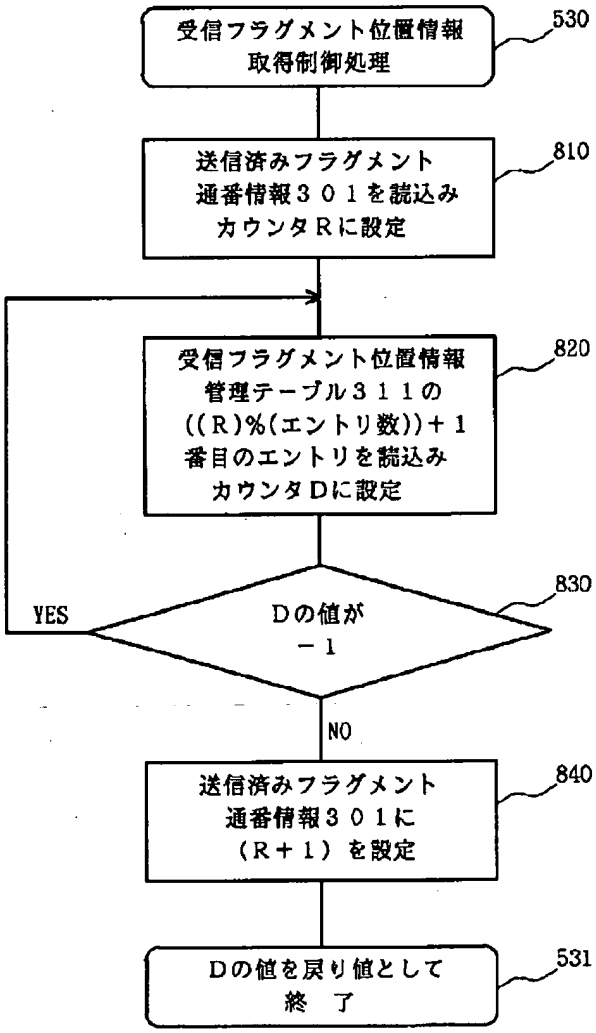
【図7】

図 7


データ送信制御処理
510
カウンタ初期化
N＝1
701
未使用フラグメント管理
制御処理手段550Aより
フラグメント取得
710
送信データ領域201のN番目の
フラグメント分のデータを
フラグメントにコピー
720
受信フラグメント位置情報取得
制御処理手段530より
受信フラグメント位置を取得
730
ネットワークアダプタ107Aを
起動してフラグメント送信
740
未使用フラグメント管理制御
処理手段550Aにフラグメントを
未使用として登録
750
カウンタ値加算
N＝N＋1
760
NO
送信データ領域201
の全てのデータの
送信処理完了
770
YES
終了
511

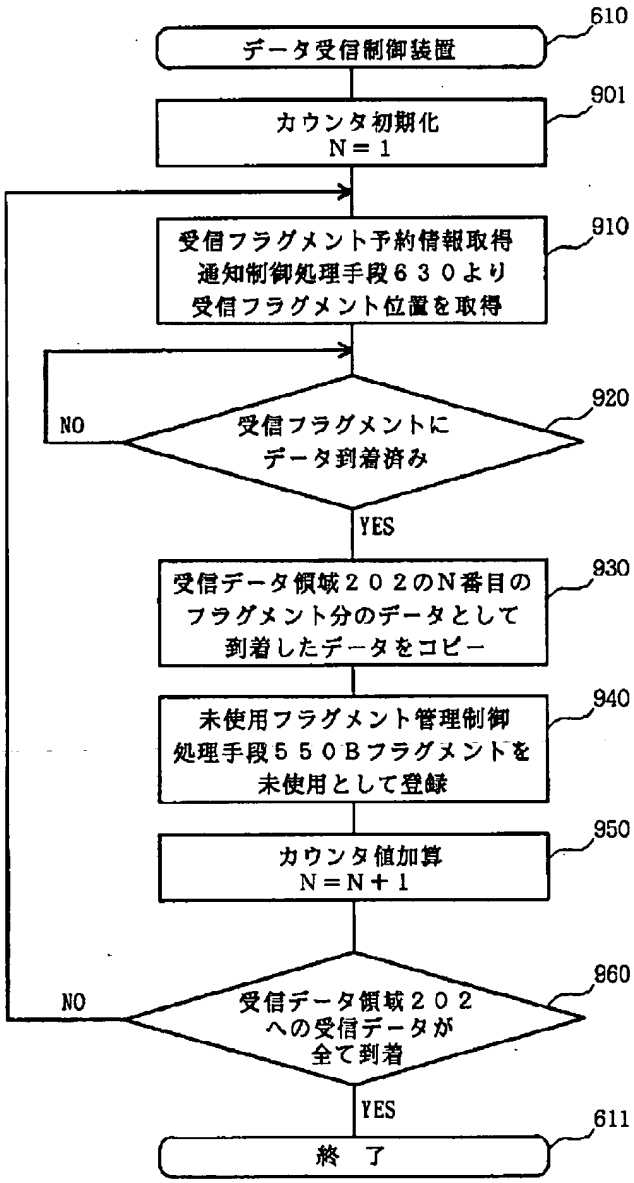
【図8】

図 8


受信フラグメント位置情報
取得制御処理
530
送信済みフラグメント
通番情報３０１を読込み
カウンタＲに設定
810
受信フラグメント位置情報
管理テーブル３１１の
((R)%(エントリ数))＋１
番目のエントリを読込み
カウンタＤに設定
820
YES
Dの値が
－１
830
NO
送信済みフラグメント
通番情報３０１に
（Ｒ＋１）を設定
840
Dの値を戻り値として
終　了
531

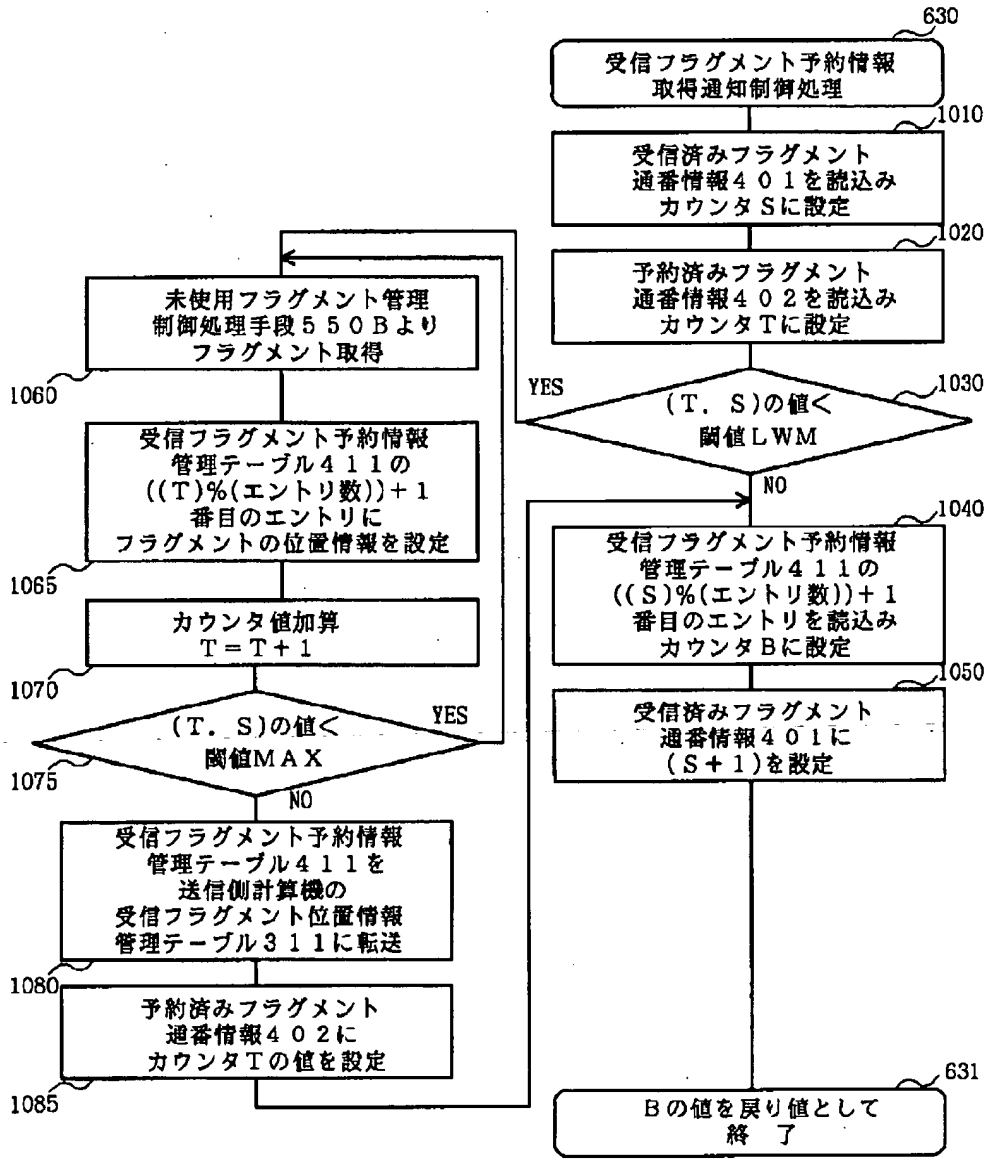
【図9】

図9


610
データ受信制御装置
901
カウンタ初期化
N＝1
910
受信フラグメント予約情報取得
通知制御処理手段６３０より
受信フラグメント位置を取得
920
NO
受信フラグメントに
データ到着済み
YES
930
受信データ領域２０２のN番目の
フラグメント分のデータとして
到着したデータをコピー
940
未使用フラグメント管理制御
処理手段５５０Bフラグメントを
未使用として登録
950
カウンタ値加算
N＝N＋1
960
NO
受信データ領域２０２
への受信データが
全て到着
YES
611
終　了

【図10】

図10



フロントページの続き

(72)発明者　坂口　明彦
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内

(72)発明者　今木　常之
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内